omicsExample1.R

aleplo

2021-11-23

# omicsExample1.R  
#  
# Code for Example 1 for R intro in course Omics Data Analysis: a simple  
# qPCR data set (not very omic, but we have to start somewhere)  
#  
# Alexander.Ploner@ki.se 2021-11-18  
  
# We assume that the file is in this directory and set it as working directory  
# Modify as required  
setwd("~/OmicsDataAnalysis/Data/")  
dat <- read.table("qPCR.txt", header = TRUE, stringsAsFactors = TRUE)  
  
# Look at the top of the data: seems ok  
head(dat)

## Sample Con DeltaCt  
## 1 Control 10 3.3687  
## 2 Control 10 3.4063  
## 3 Control 10 3.5066  
## 4 Control 2 4.5963  
## 5 Control 2 3.7704  
## 6 Control 2 3.4906

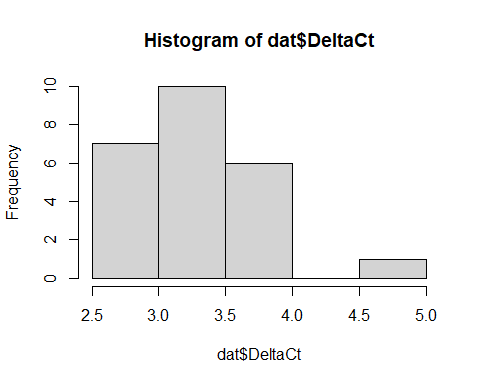
# Structure of the data: one grouping variable, four numeric variables  
str(dat)

## 'data.frame': 24 obs. of 3 variables:  
## $ Sample : Factor w/ 2 levels "Control","Treatment": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Con : num 10 10 10 2 2 2 0.4 0.4 0.4 0.08 ...  
## $ DeltaCt: num 3.37 3.41 3.51 4.6 3.77 ...

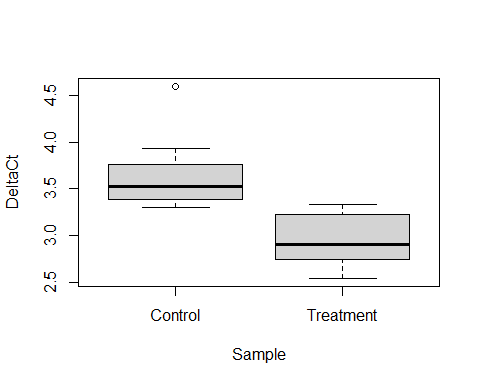
# Simple numerical summaries of the data  
summary(dat)

## Sample Con DeltaCt   
## Control :12 Min. : 0.08 Min. :2.540   
## Treatment:12 1st Qu.: 0.32 1st Qu.:2.919   
## Median : 1.20 Median :3.335   
## Mean : 3.12 Mean :3.298   
## 3rd Qu.: 4.00 3rd Qu.:3.516   
## Max. :10.00 Max. :4.596

# We want to study how DeltaCt (difference in time until detection threshold  
# between sequence of interest and reference sequence) differs between  
# treated / controls  
# Let's start by looking at the distribution of DeltaCt  
# Hm, seems like we have a fairly large value  
hist(dat$DeltaCt)



# Direct comparison: treatment vs controls  
# Looks like treated samples have lower DeltaCt,  
# ergo higher level of SOI  
boxplot(DeltaCt ~ Sample, data = dat)



# Statistical test: is DeltaCt on average the same?  
# t-test without equal variance assumption (i.e. Welch t-test)  
# Null hypothesis: same average DeltaCt in both groups  
# Based on p-value and confidence interval, we can reject this  
t.test(DeltaCt ~ Sample, data = dat)

##   
## Welch Two Sample t-test  
##   
## data: DeltaCt by Sample  
## t = 5.2577, df = 20.684, p-value = 3.429e-05  
## alternative hypothesis: true difference in means between group Control and group Treatment is not equal to 0  
## 95 percent confidence interval:  
## 0.413698 0.955952  
## sample estimates:  
## mean in group Control mean in group Treatment   
## 3.640408 2.955583

# Sensitivity analysis: does our result depend on the one  
# large values among the controls? We re-run without that  
# observation  
dat2 <- subset(dat, DeltaCt < 4)  
str(dat2) # We removed exactly one observation

## 'data.frame': 23 obs. of 3 variables:  
## $ Sample : Factor w/ 2 levels "Control","Treatment": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Con : num 10 10 10 2 2 0.4 0.4 0.4 0.08 0.08 ...  
## $ DeltaCt: num 3.37 3.41 3.51 3.77 3.49 ...

# Yes, this holds up well; the original result (to be reported)  
# is not unduly affected by just one observation  
t.test(DeltaCt ~ Sample, data = dat2)

##   
## Welch Two Sample t-test  
##   
## data: DeltaCt by Sample  
## t = 5.9708, df = 20.05, p-value = 7.644e-06  
## alternative hypothesis: true difference in means between group Control and group Treatment is not equal to 0  
## 95 percent confidence interval:  
## 0.3890666 0.8067849  
## sample estimates:  
## mean in group Control mean in group Treatment   
## 3.553509 2.955583